AMENDMENT TO THE CLAIMS

1-26. (Canceled)

27. (Cancelled)

28-30. (Cancelled)

31. (Cancelled)

32. (Cancelled)

33. (Cancelled)

34-39. (Cancelled)

40. (Cancelled)

41-45. (Cancelled)

46. (Cancelled)

47. (New)   A computer-implemented method for evaluating a word segmentation language model, comprising:

> building the word segmentation language model;

> utilizing a computer processor that is a functional component of the computer to apply the language model to a test corpus of unsegmented text different from the annotated corpus so as to provide an output indicative of words in the test corpus and a word type indication for each word, the word type indication being any one of a plurality of word type indications;

> utilizing the processor to compare the word type indication for each word in the output of the language model with predefined word type indications of words of the test corpus; and

> utilizing the processor to generate a quantitative value that represents a level of precision with which word type indications were applied in the output indicative of words in the test corpus, wherein generating comprises

generating based on a comparison of the word type indication for words in the output to the predefined word type indications.

48. (New) The method of claim 47, wherein building the word segmentation language model comprises building the word segmentation language model based on an annotated corpus.

49. (New) The method of claim 47, wherein generating the quantitative value further comprises generating a quantitative value based on a comparison of word type indications of words in the output that match predefined word type indications assigned to the same words in the test corpus.

50. (New) The method of claim 47, wherein generating the quantitative value comprises generating a quantitative value that is indicative of how frequently a word type indication in the output matches a corresponding predefined word type indication in the test corpus.

51. (New) The method of claim 47, wherein generating the quantitative value comprises generating a quantitative value that is indicative of how frequently a word type indication, assigned to a word in the output, matches a predefined word type indication assigned to a same word in the test corpus.

52. (New) The method of claim 47, wherein generating a quantitative value further comprises generating a quantitative value that represents a level of precision with which person name word type indications were applied in the output.

53. (New) The method of claim 47, wherein generating a quantitative value further comprises generating a quantitative value that represents a level of precision with which location name type indications were applied in the output.

54. (New) The method of claim 47, wherein generating a quantitative value further comprises generating a quantitative value that represents a level of precision with which organization name word type indications were applied in the output.

55. (New) The method of claim 47, wherein generating a quantitative value further comprises generating a quantitative value that represents a level of precision with which overlapping ambiguous string word type indications were applied in the output.

56. (New) The method of claim 47, wherein generating a quantitative value further comprises generating a quantitative value that represents a level of precision with which covering ambiguous string word type indications were applied in the output.

57. (New) A computer-implemented method of evaluating word segmentation models, comprising:

        utilizing a computer processor that is a function component of the computer to apply a first word segmentation model so as to segment a corpus of text into words, wherein the computer processor also assigns, to each of the words, a tag indicative of a particular one of a plurality of word types, and wherein these words and their assigned tags together signify a first output;

        applying a second word segmentation model so as to segment the corpus of text into words, wherein each word is assigned a tag, the tags being indicative one of a plurality of word types, and wherein these words and their assigned tags together signify a second output;

        generating a first set of values indicative of how effectively the first word segmentation model recognized each of the plurality of word types, wherein generating a first set of values comprises comparing the first output to a predefined indication of words identified with word types included in said plurality;

generating a second set of values indicative of how effectively the second word segmentation model recognized each of the plurality of word types, wherein generating a second set of values comprises comparing the second output to the predefined indication of words identified with word types included in said plurality; and

providing an indication of a comparison of the first set of values and the second set of values, wherein the indication of the comparison comprises a representation of effectiveness of the first word segmentation model and the second word segmentation model with respect to each of the plurality of word types.

58. (New)   The method of claim 57, wherein the first set of values is based on matches between the first output and the predefined indication, and wherein the second set of values is based on matches between the second output and the predefined indication.

59. (New)   The method of claim 57, wherein the plurality of word types includes a person name type.

60. (New)   The method of claim 57, wherein the plurality of word types includes a location type.

61. (New)   The method of claim 57, wherein the plurality of word types includes an organization name type.

62. (New)   The method of claim 57, wherein the plurality of word types includes a overlapping ambiguous string type.

63. (New)   A computer-implemented method for performing word segmentation, the method comprising:

> receiving an input of text;
>
> utilizing a computer processor that is a functional component of the computer to apply a language model so as to determine a segmentation of the text;
>
> identifying a morphologically derived word within the text; and
>
> providing an output that includes the segmentation of the text, and also includes an indication of a combination of parts that form the morphologically derived word.

64. (New) The method of claim 63, wherein the output also includes an indication of a named entity detected within the text.

65. (New) The method of claim 63, wherein the output also includes an indication of a factoid detected within the text.

66. (New) The method of claim 63, wherein the output also includes an indication a part of speech represented by the combination of parts.